DOCUMENT RESUME

ED 380 508 TM 022 875

AUTHOR	Davis, Wesley
TITLE	Alternative Assessment: Facts and Opinions.
INSTITUTION	Florida Educational Research Council, Inc., Sanibel.
PUB DATE	94
NOTE	34p.
AVAILABLE FROM	Florida Educational Research Council, Inc., P.O. Box 506, Sanibel, FL 33957 ($4; annual subscription, $15; 10% discount on 5 or more copies).
PUB TYPE	Collected Works - Serials (022)
JOURNAL CIT	Florida Educational Research Council Research Bulletin; v25 n4 p1-32 Sum 1994
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Cost Effectiveness; *Educational Assessment; Educational Change; Educational Improvement; Elementary Secondary Education; *Evaluation Methods; Literature Reviews; Norm Referenced Tests; *Opinions; Standardized Tests; *Student Evaluation; Teacher Role; *Test Construction; Test Use
IDENTIFIERS	*Alternative Assessment; Large Scale Programs

ABSTRACT

An attempt is made to separate facts from opinions based on review of a representative sample of contemporary writings on alternative assessment. A summary listing of 15 statements perceived to be factual is offered, followed by opinions of the author. These items cover: (1) the historical background and origins of alternative assessments; (2) their current intent, focus, and emphasis; (3) their technical problems and limitations; (4) the potential impact for change these procedures may have on instruction and student-teacher relationships; (5) other possible consequences of changes; (6) the expanded role of teachers in implementation; (7) the most significant contribution alternative assessment might make for students; and (8) projected cost factors. It is suggested that changing the instructional process for the better may well be the major contribution of alternative assessment. Large-scale norm-referenced standardized tests are here to stay, and cost factors may mean that alternative assessments are most useful in the individual classroom. One table summarizes facts about alternative assessment. (Contains 44 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

9570

FLORIDA EDUCATIONAL
RESEARCH COUNCIL

RESEARCH BULLETIN

ED 380 508

Alternative Assessment: Facts and Opinions

Wesley Davis
Escambia County Public Schools
Pensacola, Florida

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
C. T. Council
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 022 875

Volume 25 Summer 1994 Number 4

FLORIDA EDUCATIONAL
RESEARCH COUNCIL, INC.

RESEARCH BULLETIN

Alternative Assessment: Facts and Opinions

Wesley Davis
Escambia County Public Schools
Pensacola, Florida

Additional copies of this book may be obtained from:
FERC, Inc.
P.O. Box 506
Sanibel, Florida 33957

Individual Copies $4.00
Annual Subscription $15.00

10% Discount on an order of 5 or more.
Order of less than $20.00 must be accompanied by
a check or money order.

F.E.R.C. NOTES ON THIS BULLETIN

Differences and debate on alternative assessment, what it means and how it works seems to be a topic of great interest, especially among those whose primary responsibility is to see that it is implemented in a professionally defensible manner.

F.E.R.C. published a bulletin on this topic in the fall of 1993 and this 1994 publication certainly compliments and supplements the earlier one. Without a doubt this topic will continue to spark controversy and conversation. F.E.R.C. is pleased to bring this and other information on education to its readers.

Charlie T. Council
Executive Director

BOARD OF DIRECTORS

Lee Rowell - 1993-96
PRESIDENT
445 West Amelia Street
Orlando, FL 32801

Betty Hurlbut - 1992-95
PAST PRESIDENT
426 School Street
Sebring, FL 33870

Madhabi Banerji - 1993-96
TREASURER
7227 U.S. Highway 41
Land O'Lakes, FL 33537

Cecil Carlton - 1992-95
P.O. Box 1470
Pensacola, FL 32597

Sandra McDonald - 1992-95
40 Orange Street
St. Augustine, FL 32084

Jayne Hartman - 1992-95
2909 Delaware Avenue
Ft. Pierce, FL 34947-7299

Mary Kay Habgood - 1993-96
215 Manatee Avenue West
Bradenton, FL 34205

Rick Nations - 1993-96
2418 Hatton Street
Sarasota, FL 34327

John Hilderbrand - 1994-97
901 E. Kennedy Boulevard
P.O. Box 3408
Tampa, FL 33601-3408

John Headlee - 1994-97
3741 W. Educational Path
Lecanto, FL 32661

Tom Conner - 1994-97
P.O. Box 787
LaBelle, FL 33935

Mike Jones - 1994-97
2055 Central Avenue
Ft. Myers, FL 33901-3988

ADVISORS

Jake Beard
College of Education
Florida State University
Tallahassee, FL 32306

Carl Balado
College of Education
University of Central Florida
Orlando, FL 32816

Charles Hayes
Central Florida Community College
P.O. Box 1388
Ocala, FL 32678

Theresa Vernetsen
College of Education
University of Florida
Gainesville, FL 32611

Rodney Smith
Florida Education Center
Department of Education
325 W. Gaines Street, Suite 414
Tallahassee, FL 32399

Allen Fisher
College of Education
Florida International University
Tamiami Trail
Miami, FL 33199

Suzanne Richter
Florida Gulf Coast University
17595 S. Tamiami Trail
Suite 200
Fort Myers, FL 33908

Tom Gill
College of Education
Florida Atlantic University
Boca Raton, FL 33431

William Castine
College of Education
Florida A & M University
Tallahassee, FL 32307

Bette Soldwedel
College of Education
University of North Florida
Jacksonville, FL 32216

Phyllis NeSmith
Florida School Board
Association
P.O. Box 446
Nocatee, FL 33864

John Follman
College of Education
University of South Florida
4204 Fowler Avenue
Tampa, FL 33620

Sam Mathews
Educational Research &
Development Center
University of West Florida
Pensacola, FL 32514

Charlie T. Council
EXECUTIVE DIRECTOR
P.O. Box 506
Sanibel, FL 33957

COUNCIL MEMBERS

County	Address	Contact Person
Alachua	620 E. University Ave. Gainesville, FL 32601	Mel Lucas
Charlotte	1445 Piatti Drive Punta Gorda, FL 33948	John Wiegman
Citrus	3741 W. Educational Path Lecanto, FL 32661	John Headlee
Collier	3710 Estey Avenue Naples, FL 33942	Donn Godin
Dade	1450 N.E. 2nd Avenue Miami, FL 33132	Marilyn Neff
DeSoto	530 LaSolona Avenue Arcadia, FL 33821	Adrian H. Cline
Escambia	Post Office Box 1470 Pensacola, FL 32597	Cecil Carlton
Glades	Avenue K and 8th Street Moore Haven, FL 33471	Gary Clark
Hardee	Post Office Drawer 678 Wauchula, FL 33873	Derrel Bryan
Hendry	Post Office Box 787 LaBelle, FL 33953	Tom Conner
Hernando	919 North Broad Street Brooksville, FL 34601	Jim Knight
Highlands	426 School Street Sebring, FL 33870	Betty Hurlbut

COUNCIL MEMBERS (cont.)

County	Address	Contact Person
Hillsborough	Post Office Box 3408 Tampa, FL 33601	John Hilderbrand
Indian River	1990 25th Street Vero Beach, FL 32960	Jack McAfee
Lee	2055 Central Avenue Ft. Myers, FL 33901	Mike Jones
Manatee	215 Manatee Avenue West Bradenton, FL 34205	Mary K. Habgood
Marion	Post Office Box 670 Ocala, FL 32670	Esther Oteiza
Martin	Post Office Box 1049 Stuart, FL 33494	Deana Hughes
Nassau	1201 Atlantic Avenue Fernandina Beach, FL 32034	James T. Northey
Okeechobee	100 S.W. 5th Avenue Okeechobee, FL 33472	Danny Mullins
Orange	445 West Amelia Street Orlando, FL 32802	Lee Rowell
Pasco	7227 U.S. Highway 41 Land O'Lakes, FL 33537	Madhabi Banerji
Polk	1915 So. Floral Avenue P.O. Box 319 Bartow, FL	Mary Topping
St. Johns	40 Orange Street St. Augustine, FL 32048	Sandra McDonald

COUNCIL MEMBERS (cont.)

County	Address	Contact Person
St. Lucie	2909 Delaware Avenue Ft. Pierce, FL 34947	Jayne Hartman
Sarasota	2418 Hatton Street Sarasota, FL 33577	Rick Nations
Suwannee	224 Parshly Street Live Oak, FL 32060	Marvin Johns
Volusia	729 Loomis Avenue P.O. Box 2410 Daytona Beach, FL 32115	Cynthia Pino

AFFILIDATED MEMBERS

County	Address	Contact Person
Baker	392 So. Boulevard E. MacClenny, FL 32063	William Baker
Bay	1311 Balboa Avenue Panama City, FL 32401	Jo Ann Cox
Brevard	2700 St. Johns Melbourne, FL 32940	Dan Schuerer
Jackson	P.O. Box 5958 Marianna, FL 32447	Frank Waller
Levy	P.O. Drawer 129 Bronson, FL 32621	Ed Lovely
Leon	2757 W Pensacola Street Tallahassee, FL 32304	Shirley Eikeland
Santa Rosa	603 Canal Street Milton, FL 32570	Richard Mancini

Abstract

The present paper seeks to separate facts from opinions based upon review of a representative sample of contemporary writings on alternative assessment. A summary listing of fifteen perceived to be factual statements is offered, followed by the present writer's opinions. These items cover the historical background and origins of these assessments; their current intent, focus, and emphasis; their technical problems and limitations; the potential impact for change which these procedures may have upon both the instructional process and teacher-student relationships, the possible consequences of such changes, and the newly expanded role of classroom teachers in implementation; the likely most significant contribution which these procedures might make on behalf of students in contemporary public education; and projected cost factors. Much of the above has also been set forth in an inferential summary.

Introduction

When reading in the area of alternative/authentic/performance assessment, one experiences considerable difficulty separating fact from fiction. This makes even more interesting the use of the word "authentic" by those who feel that it somehow lends additional credence to these procedures. Is there anything less authentic about a student taking a standardized, norm-referenced achievement test? Is there not also an assessment of student performance in the use of such a test? This apparent play upon words also adds to the confusion prevalent in this endeavor. Hence, given this state of affairs, it is the intent of the present paper to attempt a separation of at least some of what appears to be fact from more obvious opinion, with the ultimate decision relative to each resting with the reader.

Discussion

The format of this paper will consist of what are perceived to be factual statements (extracted from considerable readings) followed by the present writer's opinions. As the reader may well appreciate, it is quite difficult at certain points to differentiate clearly between the two. Obviously what is offered involves perceptual judgment, and in such matters there is seldom solid agreement. Nonetheless,

positions have herein been taken; and some will say with obvious bias. To what extent these statements provide even a modest degree of clarification will, of course, be decided by the critics. Some of these folk will have read widely and will possess the necessary technical background and experiences to appreciate the current level of confusion which prevails. Others, without the benefits of either, will voice their opinions as well; and in this, somewhere, there will be either more or less sound and fury.

Fact: So-called alternative/authentic/performance assessments are not new.

Performance assessments have been around since Adam and Eve, with Biblical accounts alone being numerous. The historical roots lie in antiquity (Ward & Murray-Ward, 1993). Every actor who ever performed, every musician or vocalist, every artist, every worker for all time, and millions of others could speak to performance assessments. No serious athlete ever performed without knowing and feeling the meaning of assessment. Crowds have cheered or booed their approval or disapproval for eons.

Opinion:

The current emphasis upon the use of alternative assessments in education is but one of several "new" fads vying for center stage. There is probably no other profession known to man where fads come and go or the pendulum swings more rapidly. Common expressions like "here we go again" and "this too will pass" certainly are not without meaning or relevance. This particular fad (as presently conceived), however, appears likely to have a relatively brief history; and there are some fairly good reasons why. A number of them will become increasingly apparent as this paper develops. But apart from this, what needs to be better understood is not that public education is trying some new form of student assessment but that the real push in all of this is to reduce the present reliance upon standardized testing. Herein lies the problem. The current emphasis and directional aim of alternative/authentic/performance assessment implementations have the wrong focus. This point will be developed further immediately below.

Fact: Performance assessments have been going on since the very first teacher-student relationship.

As stated above, there is very little that's "new" about performance assessments. One hears the expression all the time: "Good teachers have been using alternative assessment procedures for centuries. The only thing new is the emphasis and the variety of forms which they take." A number of writers have made this point quite well (Aschbacher, 1993; Burnham, 1986; Haney & Madaus, 1989; "Performance assessment," 1994).

Opinion:

Perhaps the single most significant contribution which alternative assessment procedures potentially can make will be in their impact upon the teacher-student relationship. Teachers will better understand, in more specific terms, what individual students can and cannot do. They will see more clearly what impact their teaching has upon every student with whom they work. This will positively change the instructional process, and that is precisely what should be the intent, emphasis and focus of the implementation of these procedures. From that changed process should come a greater likelihood of increased individualized instruction. This approach encourages students to develop and learn more at their own pace; and, more importantly, they will tend to see themselves competing with whom they potentially can be rather than with other students in the classroom. This could positively affect the overall learning atmosphere. It could also afford the teacher a greater sense of accomplishment in having helped the student to attain a more clearly perceived potential. The impact of this process, through this specific focus and emphasis, could be a notable contribution. Indeed, public education in the mid-1990s needs all the positive contributions it can get.

Fact: Part of the current emphasis upon the use of performance assessments in education is related to the influence of business and industry through representatives on boards of accountability, student assessment, etc.

Most folks would probably agree that contemporary public education continues to face numerous crises. National test scores, international competition, numerous governmental reports, state and national conferences, etc., all tend to support this inference (Fisher, 1993a; Miller & Legg, 1993; Spady, 1992; Stiggins, 1991). School improvement is a hot topic throughout the nation (Darling-Hammond

& Wise, 1985). That representatives from business and industry are having increased input in this process would not be seriously questioned; and many would agree that this is as it should be.

Opinion:

While there is considerable agreement that performance assessments are certainly not new, the modern emphasis is indeed new and has some rather specific origins. Essentially, it grows out of vocational education, as strongly supported by business and industry. The use of work samples, as in assessing student or trainee progress, in evaluating the effectiveness of apprenticeship programs, and in judging the quality of a worker's performance, has a long and favorable history. Thinking in these terms, it is obvious that those who represent work-related interests would promote the adoption of this concept in public education. The problem with this, however, is that very much of what goes on in the typical classroom does not lend itself to these limited and discrete units of assessment. Critics, of course, say this is the very reason why the classroom has to change to become more "authentic" or real-world oriented. To some extent they have a legitimate point; but, unfortunately, many of those speaking do not have the background in assessment to appreciate the inherent problems or to see that the process simply will not work in the same manner as it does in business and industry. Trying to teach a child to read; to appreciate literature, music, the arts, philosophy, the complexities of mathematics, etc.; or to learn ethical principles and morality is not the same as teaching one to repair a toaster, paint a house, lay a sidewalk, run a drillpress, or whatever. The sincerity of the intent is seductively persuasive, but is does not alter the complexity of the task. Moreover, to assess progress toward the completion of that task is equally complex. As the essential crux of the matter, this complexity and the inherent problems which cause it are not fully understood or appreciated.

Fact: There are problems associated with alternative assessment procedures which many of us, including most classroom teachers, do not yet understand.

These problems are primarily technical in nature. Not the least of them are matters involving reliability and validity (Baker, O'Neil, & Linn, 1993; Gearhart, Herman, Baker, & Whittaker, 1993; Koretz,

McCaffrey, Klein, Bell, & Stecher, 1992a; Linn, Baker, & Dunbar, 1991; Messick, 1992). Of concern here is reliability involving both test item performance and test item scoring. In this area, in particular, both are intimately related. Since most performance assessments involve a very limited number of tasks (in many cases only one), conventional reliability estimates are either not applicable or produce results which are markedly unstable. Repeated studies have indicated (referenced in Fisher, 1993a) that a test must have a minimum of at least six items even to approach acceptable reliability estimates. Moreover, while the notion of increasing the number of items to improve test score reliability is a time-honored one, the concept seems to have little applicability to contemporary approaches to performance assessment (Stiggins, 1987, 1991; Worthen, 1993). This, or course, leaves validity blowing in the wind. So if we have questionable reliability for both test item performance and test item scoring, how does one defend as valid any assessment of such limited scope? As yet, there is no entirely satisfactory answer.

Opinion:

There appears to be very limited optimism relative to a positive answer for the above question. The technical experts have been amazingly quiet on these issues. Perhaps it's not politically feasible to leap into these areas, or maybe it is perceived as a "no win" situation. At any rate, a good many of the persons talking and writing about alternative assessments appear to be among the least able to provide the much needed answers. This would seem to be a beautiful, wide-open area for some young statistician to achieve fame, though probably not fortune. Who will break the silence?

Fact: The way some performance assessments evolve or unfold raises serious questions relative to their value as an assessment tool.

When one reads about the various programs which promote the use of performance assessments, there appears to be a gross lack of understanding of the basic principles associated with tests and measurement or commonly accepted assessment procedures (Dunbar, Koretz, & Hoover, 1991; Fisher, 1993a; Gearhart et al., 1993; Goldman, 1993; Herman, Aschbacher, & Winters, 1992; Koretz et al., 1992a; Koretz, Stecher, & Deibert, 1992b; Rothman, 1994; Shavelson, Baxter, & Pine, 1992; Shavelson et al., 1993; Vernetson, 1993; Winograd,

n.d.). There is, for instance, progressive scoring, editing, and revising of the performance or product; continual input from others in a cooperative setting; and a continual shifting of selection criteria for the performance or product to be evaluated. This is an assessment nightmare. It raises all kinds of questions about whose work is being evaluated, at what point in time, and under what conditions (Gearhart et al., 1993; Shavelson et al., 1993). Would anyone attempt to do the same thing with a student's performance on a standardized test? Absolutely not, unless that person wanted to risk losing a teaching certificate, face possible dismissal, or deal with a couple of justly irate parents. Beyond this, how will one truly know when, or if, the student has actually achieved what? Yet, some are willing, and even eager, to use these performance data to claim that students are making progress and attaining projected goals related to "authentic," real-life situations (Gordon, 1991; Miller & Seraphine, 1993; Shavelson et al., 1993; Wiggins, 1989a, 1989b).

Opinion:

This point appears to be intimately related to the anxiety which a good many classroom teachers feel relative to data from standardized testing programs. In effect, however, alternative assessment, as commonly conceived, never actually occurs. It's always an on-going process; the target constantly moves; the student is in a perpetual state of "growing" or "improving" such that no one really knows where the student is at any given point in time. This takes the pressure off of teachers, which is a subtle but primary intent of this movement; and since they do the scoring of the performances or products anyway, they have gained greater control over the assessment process. Well, unfortunately, what we really have with such an arrangement are two things: 1) more of an age-old scenario of "the fox guarding the chickens" (which education needs less of rather than more of), and 2) a teacher deciding, without external corroboration (i.e., standardized test results), how much progress students have made under his/her tutelage. How many teachers have you known who say that students made little to no progress while under their care? Self-preservation is a powerful force. It has been known to distort perceptions, even under the best of conditions, and even at the expense of students. An "external," standardized test has been only rarely so accused.

Fact: The scoring of most alternative assessment products or performances lacks objectivity.

While such things as scoring rubrics, scorer training sessions, and supervised experiences are attempts at improving scoring objectivity, it yet remains that the scoring of student performances and products is essentially subjective (Aschbacher, 1993; Costa, 1989; Gordon, 1991; Haney & Madaus, 1989; Linn et al., 1991; Worthen, 1993). There are some who tend to see this as more of an advantage rather than a disadvantage (Burnham, 1986; Gordon, 1991; Neill & Medina, 1989; Shepard, 1989; Valencia & Pearson, 1987; Wiggins, 1989b). The subjective nature of the assessment process is thought to add to the "authenticity" of the overall experience.

Opinion:

If it is true that the subjective nature of the assessment process adds to its authenticity, then why have rubrics or training sessions? Not having them would certainly seem to increase the subjectivity of the experience, that is at least for the assessor or evaluator. At the heart of this matter, of course, is the question of equity. Is the student being treated fairly when these highly subjective scoring procedures are in use? This would certainly seem to be one of the reasons why throughout the history of psychological and educational testing there has been a steady move toward increased standardization and objectivity.

Fact: The scoring of alternative assessment products or performances causes considerable concern among a good many who are trained in measurement theory.

There are several items that touch the core of this concern. Among them certainly are reliability and validity (Aschbacher, 1993; Baker et al., 1993; Dunbar et al., 1991; Gearhart et al., 1993; Koretz et al., 1992a; Linn et al., 1991; Messick, 1992); the underlying scale, which is intended to represent the actual measurement (Worthen, 1993); whether or not there is an accompanying rubric, how it is used, will it represent the skills and cognitive domain being assessed; what kinds of training, if any, the scorers have received, etc. (Dunbar et al., 1991; Fisher, 1993a, 1993b; Gearhart et al., 1993; Goldman, 1993; Miller & Seraphine, 1993; Shavelson et al., 1992; Stiggins, 1991;

Valencia, 1991; Vernetson, 1993; Wiggins, 1989a, 1989b; Winograd, n.d.). This is a limited list, both of the items of concern and of the writers who have spoken to them; but it is certainly representative.

Opinion:

Resting at the heart of these concerns is scorer reliability. There are so many factors which potentially affect the response mode of the scorer. The instability is notorious, especially when extended over time (to which the results of several state programs will currently attest). But perhaps equally important, however, is the underlying measurement scale. Let's assume, for instance, that the scale has seven points, which is quite common with these types of assessment. Writing assessment would clearly be a good example. Now if the scale has seven points, the midpoint, or "average," is said to be 3. This leaves 0-2 on one end and 4-6 on the other. So how is it used? Even without the first minute of training, the scorer should be able to operate with this scale quite comfortably—and usually does, with apparent limited concern for what the scores might mean. Basically, only **one** discrimination needs to be made. Is the student's product or performance **above** or **below** average? If below, assign a 2 (not a 1, since a 0 is almost never assigned); if above, assign a 4 (not a 5, since a 6 is only rarely assigned). This immediately allows the scorer to have a very high percent of agreement with most other scorers, who essentially (though unknowingly) are operating the same way. If the above discrimination **cannot** be made comfortably, the scorer simply assigns a 3 (for average), again permitting high scorer agreement.

On those occasions when the scorer encounters an exceptional product or performance, a score of 5 is assigned. For those who happen to score the same performance as a 4 or 6 (since discriminations at the upper end of the scale tend to be more difficult) there is once again a notably high percent of agreement. This same process, though not equally difficult, is repeated for the other end of the scale. The obvious reported outcome, across the entire scale, is a relatively high percent of agreement, which gives the appearance of credence and supports the distinct impression that the process is both reliable and valid.

As one might imagine, classroom teachers tend to be quite comfortable with this type of assessment, in large part because of a long history of having used a 5-point grading scale. With by far the majority having had little to no training in measurement theory, they

see no problem in using a scale of such limited range. For them, so long as there is rater "agreement," that's all that really matters. Well, "agreement" is indeed important; but it is not the whole story.

There is yet another measurement issue here of perhaps even greater importance. It has to do with the type of measurement involved, that is norm-referenced versus criterion-referenced. Most raters probably score the product or performance with a cognitive frame of reference reflecting normative assessment. Cognitively, they focus upon the hypothetical "average and score all products or performances relative to that point of reference. This is in direct conflict with the intent of the rubric, which is built upon the notion of criterion-referenced measurement. Specifically, each segment of the rubric is intended to have a distinct set of criteria which supposedly delineate what the scoring process is all about. This is certainly the essence of the training sessions for potential scorers.

Once the (assumed to be "criterion-referenced") scores have been obtained, however, they tend to be treated as though they represented norm-referenced measurement. So cognitively, for the scorers, they are right back to their point of origin. This is confusion to say the least. The bottom line is: These types of scores are very different; they are treated differently in statistical analyses; and they are interpreted differently. The shape of the frequency distributions alone dramatically illustrates the point. With different underlying assumptions; and specific limitations as to commonly accepted data treatment or analyses; subjectively derived raw scores from performance assessments cannot, and in the present writer's judgment will not, displace the more stable, objective scores provided by broad-scale, norm-referenced assessments. Sound reason won't permit it, apart from of the rather consistent and stable posture of the supporting public.

Fact: Any assessment procedure is intended to reflect the underlying measurement scale which seeks to represent or embody the meaning or essence of that assessment.

What kind of measurement scale is commonly represented by the types of performance assessments currently being promoted? Most of these scales with which the present writer is familiar (Fisher, 1993a; Herman et al., 1992; Stiggins, 1987) are assumed to be ordinal and to have a maximum score range, as indicated above, from about five to seven points. By contrast, the typical standardized test has an

underlying statistical scale which is defensible as being somewhere between ordinal and interval measurement and has a maximum score range anywhere from about twenty to a hundred or more points. When considering the total battery, many of these instruments have scales representing several hundred points. This permits, or course, the assessment of many concepts; it greatly increases the stability of the scores; and it potentially impacts validity in a positive manner. These are the very issues which tend to impune the wholesale, expansive use of performance assessments as currently conceived.

Opinion:

The majority of persons currently promoting the use of alternative/authentic/performance assessments are neither specifically trained nor generally informed relative to the measurement scales which underline these assessment procedures. When asked questions in this area, they either bow out or quickly defer to someone else who is perceived by them to be better informed. Classroom teachers are, of course, familiar with the common grading scale; and most of them have a general understanding that each grade represents some point on an underlying measurement scale. They tend not, however, to be familiar with the notion of different levels of measurement or how data generated from these different scales are subject to restraints in terms of interpretation, processing, generalizability, etc. For a good many of them, numbers are numbers; and they are prone to believe that these numbers can be manipulated mathematically at will. So why all the fuss over what the scores mean? Anybody knows what an **average** is; so the student's performance is either average or not average. If the performance is rated as above average, all is fine; if below, then there is more work to be done. That's what performance assessment is all about, right? Well, maybe yes; and maybe no.

If one is talking about a particular student relative to an **internal** standard; and that student is attempting to achieve his/her own standard, then the answer is probably yes. The teacher is only seeking to facilitate the process. However, if one is talking about some broader standard, a group norm, for instance, then the answer is no. The requirements of measurement are totally different. They are far more stringent, and the limitations relative to applicability are far greater. Most current performance assessments do not meet these criteria. Moreover, the prospects for their notable improvement are

not too promising. Perhaps it is in large part for these reasons that the measurement experts have been amazingly silent. That this is true is distressing, to say the least. Moreover, the present paper is, if nothing else, an appeal for them to speak out.

Fact: The limited standardization of alternative assessment procedures restricts generalization of the results.

Some would argue that generalization of the results from alternative assessment procedures is neither a high priority nor a primary goal of the process (Anrig, 1993; Arter & Spandel, 1992; Baker, 1989; Herman, 1992). Instead, the uniqueness of the experience, for a given student, is touted as a highly desirable feature (Gordon, 1991). The contents of a portfolio, for instance, are unique to a specific student. In no way are they intended to reflect the work, skills, strengths, weaknesses, etc. , of any other student. In a good many situations, comparisons are not even encouraged. There is no specific intent to generalize the assessment or evaluation of these contents to other individuals (Burnham, 1986).

Opinion:

While individuality and uniqueness are indeed desirable when focusing upon a specific student, these characteristics become barriers when one faces the need for data to reflect the broader impact of the educational endeavor. The fundamental question here touches not only upon the intent of the assessment process itself but also upon the ultimate value of the results. While the results for a specific student may be thought to have distinctive value, that value markedly diminishes when one attempts to make comparisons between individual students or across groups of students.

Educational leaders are clearly interested in how the overall program affects students as groups, not just as individuals. Accountability requirements alone demand as much. The reporting of school, district, state, and national educational data is clearly mandated in school board rules, as well as in both state and federal legislation. The accountability stipulations are precise; and they are indeed in the best interests of students, the system, and the supporting public. Without data from standardized assessment instruments, these mandates cannot be met. Alternative assessment procedures, as currently conceived, will not meet this need.

Some would offer students' writing performance as an exception. "Everyone writes on the same prompt; that's standardization." Well, not exactly. While the stimulus may appear to be the same, the experience certainly is not. But beyond this, there is the problem of the extremely limited sample of the respondent's behavior. A one-item test simply does not provide data having any semblance of either stability or validity in the conventional sense (contrary to the belief of those who support state writing programs as being useful for accountability purposes). Without satisfying these basic requirements, the data lack "authentic" value.

Fact: Data generated from most performance assessments do not lend themselves to aggregation into larger units.

This has particular relevance when one considers how these data might normally be used. By far the majority of current large-scale assessment programs have as their basic intent the acquisition of data to be used in program evaluation, curricular decision-making, the distribution of funds, accountability, public reporting, etc. (Anrig, 1993; Costa, 1989; Darling-Hammond & Wise, 1985; Fisher, 1993a, 1993b; Mehrens, 1992; Rothman, 1994; Worthen, 1993). Data from present proposals for performance assessments simply cannot be aggregated and used in this manner. Their use is limited principally to acquiring new insights into the teacher-student relationship and to the periodic evaluation of student progress; and this is precisely as it should be.

Opinion:

Alternative assessment data, specifically portfolio contents, for instance, represent a significant limitation when one considers accountability issues or attempts to generalize inferences drawn from them to some larger group (which was never their intent). Thus, the restrictions relative to their use at the school, district, state, and national levels become immediately apparent and a matter of even broader concern. In this single respect alone, there again appears to be very little likelihood that these assessment procedures will displace, even to a limited extent, current external (nationally-normed, standardized) testing programs.

Obviously this latter statement (repeated a third time) reflects a rather strong opinion; and there are quite a few who would neither

concur nor support it. At present at least, this writer is unaware of any situation where the introduction of performance assessments has totally displaced contemporary standardized testing programs. The overriding opinion here is that the public outcry, were such attempted, would not permit it. Parents want to see standardized test scores; they support external evaluations of student progress; and they are not going to be completely comfortable with performance assessments alone. As a supplemental procedure, however, reflecting individual student performance, they will welcome it and support it.

Fact: Quite a few groups are moving forward with the implementation of alternative assessment procedures even though the majority of the inherent problems have not yet been solved.

One prime example of this point is the more recent test development activities of the group representing the National Assessment of Educational Progress (Fisher, 1993a). For better than two years now, performance items have been used by the NAEP group. The reviews on these activities have been mixed, which might have been anticipated. More recent reports (Rothman, 1994) indicate that some who support and promote such plans intend to move more slowly with future ventures in this area.

A number of states are also pushing forward in developing plans for alternative assessment ("Alternative assessments," n.d.; Fisher, 1993a, 1993b; Hildebrand & Blackman, 1994; Koretz et al., 1992a, 1992b; Winograd, n.d.). For some, state statutes mandate such movement; for others, various commissions and advisory boards are making these recommendations. Some view these procedures as the wave of the future and are including them in their vision statements (Fisher, 1993a, 1993b). Again, the reactions have been mixed.

Opinion:

The obvious question seems to be: Are these efforts to move forward in this area, at whatever speed, the vision of genius or something quite opposite? How one attempts to answer depends a lot upon whom one reads or with whom one talks. Whatever one's personal posture, there clearly needs to be some serious reality testing in this area. Part of this has to be a clarification of the fundamental purposes fueling this movement. An obvious one is an

attempt to curtail notably the use of standardized testing, which is simply not likely to happen. Some increase in the classroom teacher's control over the assessment process likely will occur; but, at best, this will be a compromised appeasement having a potentially positive outcome. The perceived gains will be short-lived, however; because the current emphasis upon this new fad will, in the opinion of the present writer, tend to subside rather quickly. It will happen in spite of the efforts of the Lauren Resnicks of the world, who manage to attract an awful lot of money to support their projects. Indeed, though not yet herein mentioned, it would take a great deal of money to fund these alternative assessment procedures; and cost is definitely a major factor.

Fact: The current most prevalent form of alternative assessment is the portfolio, with by far the largest entry being writing samples.

In essence, a portfolio is a folder containing performance products. Its use in performance assessment is not new, having been around for quite a long time in art groups, business affairs, and industry. Its use in education, however, is fairly new; though there is considerable variance as to how it is used, what products are included, and how its contents are scored (Arter & Spandel, 1992, Belanoff & Dickson, 1991; Burnham, 1986; Gearhart et al., 1993; Koretz et al., 1992a, 1992b; Valencia, 1991; Winograd, n.d.; Winograd & Jones, 1992). Some score these products in a holistic manner, while others score individual items within the portfolio. There is no clear-cut agreement here; and the scores tend to be unstable, resulting in notable variance. Attempts to control this instability have led to the heavy reliance upon writing samples. These tasks have the appearance of being somewhat standardized; English teachers, in particular, have an extended history of having scored students themes; and the use of a formalized rubric facilitates to some extent, scorer agreement (Herman et al., 1992; Rothman, 1994).

Opinion:

There is probably less generalized concert over the use of portfolios in student assessment than any other alternative method. Classroom teachers seem to be quite comfortable with the process. This is thought to be true for at least two reasons: 1) it is one clear area where teachers have considerable control over the assessment pro-

cess; and 2) they are quite comfortable with the method of scoring, in part because it mimics the grading scale and is global enough not to force fine discriminations (which tend to raise anxiety levels). Unfortunately, the underlying measurement scale is so gross that it affords little information of any "authentic" or real value; and it is far too easily manipulated, as was pointed out in earlier paragraphs. Moreover, attempts to aggregate these data for larger assessment purposes currently hold little promise. There are, of course, groups doing just that in a willy-nilly fashion without any apparent concern for the limitations of the measurement scale, with at best only a courtesy gesture toward standardization, and an apparent insensitivity to the need to generalize whatever inferences the data might afford. The outcomes thus far have left much to be desired. This has not, however, dampened spirit or deterred the movement. "It's onward and upward, damned be the data."

Fact: A substantial number of classroom teachers believe that moving to greater use of alternative assessment procedures will permit them to regain control of the student evaluation process.

A good many teachers feel that they have lost control of the student assessment process and that persons outside the classroom are making far too many of the decisions related to achievement testing and general data collection (Anrig, 1993; Arter & Spandel, 1992; Aschbacher, 1993; Burnham, 1986; Costa, 1989; Gordon, 1991; Haney & Madaus, 1989; Neill & Medina, 1989; Shepard, 1989; Valencia & Pearson, 1987; Wiggins, 1989b). While these teachers recognize that much of this process is aimed at program evaluation and reporting to the supporting public (Bake, 1989; Darling-Hammond & Wise, 1985; Herman, 1992; Mehrens, 1992; "Performance assessment," 1994), they do not feel a sense of "ownership" or that they have adequate input in this area. For them, any form of "external" assessment is high-stakes; and they often feel victimized by the process.

Opinion:

A good many classroom teachers are tired of feeling "victimized" by "external" testing programs and would like to remove or at least to minimize this perceived threat; and they are receiving a lot of encouragement and support from their national organization in this

endeavor. Few of them believe, however, that standardized testing will completely go away; and some of them know that it will not. They recognize that the public will continue to demand it. Again, the bottom line is: They want more control over student assessment (something which they never really lost), and they see these "newer" forms as a means of regaining it. But perhaps most important and fundamental, as a dynamic or driving force, intense feelings relative to security and self-worth are perceived to be at work here; and these particular feelings are powerfully strong motivators.

Fact: The costs associated with alternative assessment procedures are considerably greater than are those commonly assigned to more traditional or conventional assessment.

Most conventional assessment procedures today tend to range in cost from about $2.00 to $5.00 per student. This depends, of course, upon what items one includes and the extent of the procedure. Obviously these figures do not include fixed or ongoing costs but refer mainly to materials replacement and limited processing. By contrast, cost estimates related to alternative assessment procedures range from a low of about $10.00 up to $50.00 or higher per student (Rothman, 1994). The magnitude of this difference is indeed impressive and for a good many folk understandably distressing.

Opinion:

Those who clamor to dump standardized testing in favor of performance assessment (andthere are a good many already referred to in this paper) are not seeing the whole picture; and they certainly have not given sufficient consideration to the costs involved. Those who say "the tests don't measure what we teach," "there is litle to no curricular match," "standardized tests don't change anything," and the like will fall silent when faced with the reality of the costs associated with these so-called "authentic" assessment procedures. One obvious reality is that as more funds are expended for assessment, the less there are for salaries. And the quickening silence was deafening. Cost factors alone will do as much as anything else to end the present emphasis upon these "newer" forms of assessment.

Fact: There is an inverse relationship between knowledge gained about alternative assessment procedures and the probability that these procedures will displace conventional, standardized testing.

When one reads extensively in the area of alternative assessment (which the present writer strongly encourages for those responsible for decision-making in student assessment), there is an increased awareness of the significance and magnitude of the numerous problems and unanswered questions which prevail. Those who have now written in this area for several years (in particular: Baker, 1989; Burnham, 1986; Costa, 1989; Darling-Hammond & Wise, 1985; Haney & Madaus, 1989; Neill & Medina, 1989; Shepard, 1989; Stiggins, 1987; Valencia & Pearson, 1987; Wiggins, 1989a, 1989b), as well as those who have written more recently (specifically: Anrig, 1993; Arter & Spandel, 1992; Aschbacher, 1993; Baker et al., 1993; Belanoff Gearhart et al., 1993; Goldman, 1993; Gordon, 1991; Herman, 1992; Herman et al., 1992; Koretz et al., 1992a, 1992b; Linn et al., 1991; Mehrens, 1992; Messick, 1992; Miller & Legg, 1993; Miller & Seraphine, 1993; Rothman, 1994; Shavelson et al., 1992; Shavelson, 1993; Spady, 1992; Stiggins, 1991; Valencia, 1991; Vernetson, 1993; Ward & Murray-Ward, 1993; Winograd, n.d.; Winograd & Jones, 1992; Worthen, 1993), have all spoken, in one manner or another, to these problems and issues. Having read these authors, one comes to a greater appreciation for the seriousness with which these problems, issues, and unanswered questions have been struggled with throughout the history of psychological testing and educational assessment. The emphasis placed upon the need for objectivity, standardization, reliable scoring, equity, validation, generalizable results, large-group processing, economical procedures, cognitive domain clarification, and, yes, even authenticity (if nothing more than functional literacy), has indeed been impressive. An evolutionary process has clearly been at work, and the end products reflect both form and function having survived the rigors of time. In no sense have these assessment instruments come to where they are today by chance, accident, or the fickle hand of fate. To think otherwise is to risk appearing ill-informed.

Opinion:

While some authors have spoken more clearly and precisely than others, there yet remains considerable confusion and misunderstanding in this important area. Unfortunately, those who might more distinctly be classified as measurement experts have remained silent for too long a time. This has added to the confusion; and it has left a void (that yet remains) which has encouraged others less informed in this critical area to rush in to attempt to fill it. For a good many, the offerings have been weak, noninformative, and reflective of a very limited understanding of the fundamental problems inherent in these so-called alternative/authentic/performance assessments. In education, in particular, the primary emphasis has been upon relieving the stress of the classroom teacher and, through this process, regaining control over individual student assessment. In response to this, the present writer would say but two things: 1) the classroom teacher has not lost control over the process of individual student evaluation and 2) alternative assessment procedures, of whatever form or function, will not remove stress from the teacher. As a matter of fact, they will serve more to increase that stress. Who do you suppose will grade all of these products and performances, and then have to deal with the consequences of the decisions made? Not some mechanical monster called a computer/scanner, for sure. Classroom teachers will do it; and they will bemoan the day they ever agreed to it. Once this realization sinks in and becomes an individual reality, they will, as a group, become among the strongest and most vocal supporters of group-administered, standardized assessment instruments (which never left the scene anyway). This is the reality, and it is as "authentic" as it's going to get.

Summary and Conclusions

The present paper has attempted to set forth facts and opinions (summarized in Table 1) based upon inferences drawn from readings on alternative assessment; to identify and clarify further the problems, concerns, and issues which yet prevail; to offer a defense for the current use of norm-referenced, standardized testing in student assessment; and to support the notion that alternative assessment procedures have the potential to alter significantly and positively what happens to the instructional process in the typical classroom, saying, in particular, that changing that process may well be their major contribution.

It may be concluded that much of what one reads in support of the expanded use of alternative/authentic/performance assessments is little more than a testimonial on their behalf, and it is very much like a product endorsement with little to no supporting data; that the silence of the technical experts in measurement theory is engulfing, with the absence of their voices only compounding the confusion, which is both unfortunate and somewhat distressing; that in spite of the apparent attempts of some to ignore the technical problems inherent in the use of these procedures, those problems will neither spontaneously solve themselves nor go away, and their persistence will significantly impair any defensible potential which these procedures may either promise or afford; that the greatest promise of these procedures appears to be in the potential impact which they may have upon both the teacher-student relationship and the instructional process, though, unfortunately, the evaluative or assessment components of that impact will not extend or generalize beyond the immediate classroom; that cost factors alone will play a major role in the early demise of any proposed or projected expansion of these procedures; and that large-scale, norm-referenced, standardized testing programs are here to stay and will not be seriously threatened by this movement. As this process unfolds, the classroom teacher will come to a greater appreciation for these more objective, "external" assessment programs. Many of them will also gain new insights into the complexities inherent in any effective evaluation or assessment of student products and performances, be that assessment "external" or "internal"; and these new insights just may, ultimately, help them to become better teachers. The real focus in this, as always, should be upon the teacher-student relationship and the instruc-

tional process; and when this happens, the results of assessment, of whatever form or format, and any perceived consequences of these results, almost always seem to have a rather natural, "authentic" way of taking care of themselves.

Table 1. A Summary Listing of What Are Believed To Be Current Facts

About Alternative Assessment

Fact: So-called alternative/authentic/performance assessments are not new.

Fact: Performance assessments have been going on since the very first teacher-student relationship.

Fact: Part of the current emphasis upon the use of performance assessments in education is related to the influence of business and industry through representatives on boards of accountability, student assessment, etc.

Fact: There are problems associated with alternative assessment procedures which many of us, including most classroom teachers, do not yet understand.

Fact: The way some performance assessments evolve or unfold raises serious questions relative to their value as an assessment tool.

Fact: The scoring of most alternative assessment products or performances lacks objectivity.

Fact: The scoring of alternative assessment products or performances causes considerable concern among a good many who are trained in measurement theory.

Fact: Any assessment procedure is intended to reflect the underlying measurement scale which seeks to represent or embody the meaning or essence of that assessment.

Fact: The limited standardization of alternative assessment procedures restricts generalization of the results.

Fact: Data generated from most performance assessments do not lend themselves to aggregation into larger units.

Fact: Quite a few groups are moving forward with the implementation of alternative assessment procedures even though the majority of the inherent problems have not yet been solved.

Fact: The current most prevalent form of alternative assessment is the portfolio, with by far the largest entry being writing samples.

Fact: A substantial number of classroom teachers believe that moving to greater use of alternative assessment procedures will permit them to regain control of the student evaluation process.

Fact: The costs associated with alternative assessment procedures are considerably greater than are those commonly assigned to more traditional or conventional assessments.

Fact: There is an inverse relationship between knowledge gained about alternative assessment procedures and the probability that these procedures will displace conventional, standardized testing.

REFERENCES

Alternative assessments in practice database (n.d.). Los Angeles: CRESST.

Anrig, Gregory A. (1993, December). One size doesn't fit all in assessment. *The School Administrator, 50*(11), 48.

Arter, J.A., & Spandel, V. (1992, Spring). Using portfolios of student work in instruction and assessment. *Educational Measurement: Issues and Practice, 11*(1), 36-44.

Aschbacher, Pamela R. (1993). Issues in innovative assessment for classroom practice: Barriers and facilitators. *CSE Technical Report 359*. Los Angeles, CA: CRESST/CSE.

Baker, E.L. (1989). Mandated tests: Educational reform or quality indicator? In B.R. Gifford (Ed.), *Test Policy and Test Performance: Education, Language, and Culture* (pp. 3-23). Norwell, MA: Kluwer.

Baker, E.L., O'Neil, H.F., & Linn, R.L. (1993). Policy and validity prospects for performance-based assessment. *American Psychologist, 48*, 1210-1218.

Belanoff, P., & Dickson, M. (Eds.). (1991). *Portfolios: Process and product*. Portsmouth, NH: Boynton/Cook.

Burnham, C. (1986). Portfolio evaluation: Room to breathe and grow. In C. Bridges (Ed.), *Training the Teacher of College Composition*. Urbana, IL: National Council of Teachers of English.

Costa, A.L. (1989). Reassessing assessment. *Educational Leadership, 46*(2), 2-3.

Darling-Hammond, L., & Wise, A.E. (1985). Beyond standardization: State standards and school improvement. *The Elementary School Journal, 85*, 315-336.

Dunbar, S., Koretz, D., & Hoover, H. (1991). Quality control in the development and use of performance assessments. *Applied Measurement in Education, 4*(4), 289-303.

Fisher, Thomas H. (1993a, March). Alternative assessment: National overview and the Florida scene (Unpublished lecture to administrators, support staff, and teachers of the Escambia County School District, Pensacola, FL).

Fisher, Thomas H. (1993b, Fall). Perspectives on alternative assessment: What's happening nationally? In Theresa Vernetson (Ed.), *Florida Educational Research Council Research Bulletin, 25*(1), 11-16.

Gearhart, Maryl, Herman, Joan L., Baker, Eva L., & Whittaker, Andrea K. (1993). Whose work is it? A question for the validity of large-scale portfolio assessment. *CSE Technical Report 363*. Los Angeles, CA: CRESST/CSE.

Goldman, Jay P. (1993, December). Newfangled assessment: Is it adequate and credible? *The School Administrator, 50*(11), 8-31.

Gordon, D. (1991). Implications for diversity in human characteristics for authentic assessment. *CSE Technical Report 341*. Los Angeles, CA: CRESST/CSE.

Haney, W., & Madaus, G. (1989). Searching for alternatives to standardized tests: Whys, whats, and whithers. *Phi Delta Kappan, 70*, 683-687.

Herman, Joan L. (1992). Accountability and alternative assessment. *CSE Technical Report 348*. Los Angeles, CA: CRESST/CSE.

Herman, Joan L., Aschbacher, Pamela R., & Winters, Lynn (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Hildebrand, John, & Blackman, Joe (1994, January). Proposal for development of an assessment design. Tampa, FL: Hillsborough County Public Schools and The University of South Florida.

Koretz, D., McCaffrey, D., Klein, S., Bell, R., & Stecher, B. (1992a). *The reliability of scores from the 1992 Vermont portfolio assessment program*. Los Angeles, CA: The RAND Corp.

Koretz, D., Stecher, B., & Deibert, E. (1992b). The Vermont portfolio assessment program: Interim report on implementation and impact, 1991-92 school year. Los Angeles: CRESST/CSE.

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*, 15-21.

Mehrens, W.A. (1992). Using performance assessment for accountability purposes. *Educational Measurement: Issues and Practice, 11*(1), 3-9, 20.

Messick, S. (1992). The evidential and consequential validity for performance-based assessment (Unpublished paper presented at the annual meeting of AERA, San Francisco).

Miller, M.D., & Legg, S.M. (1993). Alternative assessment in a high-stakes environment. *Educational Measurement: Issues and Practice, 12*(1), 9-15.

Miller, M.D., & Seraphine, A.E. (1993). Can test scores remain authentic when teaching to the test? *Educational Assessment, 1*, 119-129.

Neill, D., & Medina, N. (1989, May). Standardized testing: Harmful to educational health. *Phi Delta Kappan, 69*, 688-697.

Performance assessment for the classroom (1994, Winter). In *Florida Newsletter* (3), Norcross, GA: CTB/McGraw-Hill.

Rothman, Robert (1994, Winter). Assessment questions: Equity answers/proceedings of the 1993 CRESST conference. *Evaluation Comment*, Los Angeles: CRESST/CSE.

Shavelson, R.J., Baxter, G.P., & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. *Educational Researcher, 21*(5), 23-27.

Shavelson, R.J. et al. (1993). Sampling variability of performance assessments. *CSE Technical Report 361*. Los Angeles, CA: CRESST/CSE.

Shepard, L. (1989, April). Why we need better assessments. *Educational Leadership, 46*(7), 4-9.

Spady, W. (1992, May). It's time to take a close look at outcome-based education. *NASP Communique*.

Stiggins, R.J. (1987). Facing the challenges of a new era of educational assessments. *Applied Measurement in Education, 4*(4), 263-273.

Valencia, S., & Pearson, R.P. (1987). Reading assessment: Time for a change. *The Reading Teacher, 40*,726-732.

Valencia, S.W. (1991). Portfolios: Panacea or pandora's box? In F. Finch (ed.), *Educational Performance Assessment* (pp. 33-46). Seattle: Riverside Publishing Company.

Vernetson, Theresa (Ed.) (1993, Fall). Selected papers from the spring 1993 Breivogel conference at the University of Florida on alternative/portfolio assessment. *Florida Educational Research Council Research Bulletin, 25*(1), 1-192.

Ward, Annie W., & Murray-Ward, Mildred (1993, Fall). Historical roots of current practice in educational testing. In Theresa Vernetson (Ed.), *Florida Educational Research Council Research Bulletin, 25*(1), 55-65.

Wiggins, G. (1989a). Teaching to the (authentic) test. *Educational Leadership, 46*(7), 41-47.

Wiggins, G. (1989b, May). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan, 69*, 703-713.

Winograd, Peter (n.d.). *The promises and perils of performance assessment*. Lexington, KY: University of Kentucky.

Winograd, P., & Jones, D. (1992). The use of portfolios in performance assessment. *New Directions in Education Reform, 1*, 37-50.

Worthen, B.R. (1993, February). Critical issues that will determine the future of alternative assessment. *Phi Delta Kappan, 74*(2), 444-454.